Electronic Version
Stylesheet Version v1.1.1

Description

[METHOD FOR FABRICATING LOCALLY STRAINED CHANNEL]

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part of a prior application Ser. No. 10/064,705, filed August 08, 2002, which is incorporated herein by reference.

BACKGROUND OF INVENTION

[0002] Field of Invention

[0003] The present invention relates to a manufacture method for a semiconductor device. More particularly, the present invention relates to a fabrication method for a semiconductor device with a locally strained channel.

[0004] Description of Related Art

[0005] The metal-oxide-semiconductor (MOS) transistor is an important device applied in very-large-integrated-circuits such as microprocessors and semiconductor memories. In addition to a gate oxide layer and a conductive gate

structure, the MOSFET transistor further includes a source/drain region having a conductivity type opposite to that of the substrate.

[0006] Fig. 1 illustrates the cross-sectional view of a prior-art semiconductor device structure. As shown in Fig. 1, the prior art structure comprises a substrate 100, a gate oxide layer 102, a gate structure 104 and a source/drain region 106. The gate oxide layer 102 is disposed on the surface of a portion of the substrate 100, while the gate structure 104 is disposed on the gate oxide layer 102. The source/drain region 106 is arranged at both sides of the gate structure 104 in the substrate 100.

[0007] To enhance the device performances, a silicon layer, which has received a tensile strain, is sometimes used as a channel. Strained silicon layer is the result of biaxial tensile stress induced in silicon grown on a substrate formed with a material whose lattice constant is greater than that of silicon, for example a silicon germanium layer. Strained silicon is shown to modify the electron or hole mobility of the silicon layer, which in turns affect the performance of a device. Larger enhancements of electron or hole mobility require a higher amount of strain.

[0008] Moreover, to increase the number of components per IC

chip, the device dimensions must be scaled down. As the device dimension continues to shrink, the source/drain regions also become smaller. The smaller the source/drain region is, the higher is the resistance, thus reducing the current of the device and inducing over loading. However, if the junction depth of the source/drain region is increased to improve the aforementioned problems, new issues, including short channel effects and junction leakage, can arise. On the other hand, if heavy dosage implantation is used to reduce the resistance, solid solubility limitation may hamper application of heavy dosage implantation for forming shallow junction for the source/drain regions. In the prior art, the shallow junction source/drain region is coupled with shrunk spacers for preventing short channel effects. However, the resulted silicided shallow junction may lead to unacceptable junction leakage and further increase the junction resistance.

SUMMARY OF INVENTION

[0009] The invention provides a manufacturing method for a semiconductor device, wherein the channel is locally strained to modify the carrier transport properties.

[0010] The invention further provides a manufacturing method for a semiconductor device for decreasing the resistance

of the source/drain region.

[0011] Accordingly, the present invention provides a manufacturing method for a semiconductor device, which is applicable for forming a strained channel. Because the gate structure is under a compressive stress, tensile strain is introduced to the channel.

[0012] The invention further provides a manufacturing method for a semiconductor device, which is applicable for forming shallow junction source/drain regions. Because the conductivity of the source/drain region is increased, the source/drain region can adopt shallow junctions to prevent short channel effect and junction leakage.

[0013] As embodied and broadly described herein, the invention provides a method for forming a semiconductor device structure, which includes forming a silicon germanium (SiGe) layer on a substrate and a strained silicon layer on the SiGe layer. A gate oxide layer is further formed on the strained silicon layer and a gate structure is formed on the gate oxide layer. A doping step is then performed to form a lightly doped source/drain region in the strained silicon layer at both sides of the gate structure. Subsequent to forming a spacer on a sidewall of the gate structure, the gate structure and the strained silicon layer are heavily

doped with n-type dopants to form an n-type gate structure and a heavily doped drain region in the strained silicon layer. Alternatively, the heavily doped drain region is formed in the strained silicon layer and the SiGe layer at both sides of the gate structure. A cap layer is further formed over the substrate. Subsequent to a cap annealing process, the cap layer is removed.

[0014] By growing a silicon layer on a SiGe layer, the channel at the strained silicon layer is under biaxial tensile strain. Further with the cap-annealing process, a compressive stress is generated on the n-type gate structure, which in turns introduces tensile strain to the strained silicon layer. The tensile strain of the strained silicon layer enhances the horizontal mobility of electrons in the channel. As a result, the operating speed of a device can also be increased.

[0015] For the manufacturing method of a semiconductor device of the present invention, because the source/drain region is located either in the strained silicon layer or in both the strained silicon layer and the SiGe layer with higher conductivity, the resistance of the source/drain region is effectively reduced.

[0016] Further, since the resistance of the source/drain region is

effectively reduced, the source/drain region can be formed as shallow junctions, thus avoiding short channel effect and junction leakage.

[0017] For the semiconductor device of the present invention and the manufacturing method thereof, both the performance and the reliability of the device are improved.

[0018] It is to be understood that both the foregoing general description and the following detailed description are exemplary, and are intended to provide further explanation of the invention as claimed.

Brief Description of Drawings

[0019] The accompanying drawings are included to provide a further understanding of the invention, and are incorporated in and constitute a part of this specification. The drawings illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention.

[0020] Fig. 1 is a cross-sectional view of a prior-art semiconductor device structure.

[0021] Figs. 2A-2G are schematic, cross-sectional diagrams to illustrate the process flow for the manufacturing of a semiconductor device according to one aspect of this invention.

[0022] Fig 3A is cross-sectional view of a semiconductor device of the present invention with a lightly doped region formed occupying the entire thickness of the strained silicon.

[0023] Figs. 3B and Fig. 3C are cross-sectional views of a semiconductor device structure of the present invention present invention with the heavily doped drain region formed in the strained layer and in the silicon germanium layer, respectively.

Detailed Description

[0024] Figs. 2A-2G are schematic, cross-sectional diagrams to illustrate the process flow for the manufacturing of a semiconductor device according to one aspect of this invention.

[0025] Referring to Fig. 2A, a substrate 200 is provided and a silicon germanium ($Si_{1-x}Ge_x$; SiGe) layer 202 is formed on the surface of the substrate 200. A strained silicon layer 204 is formed on the SiGe layer 202. The strained silicon layer 204 has a thickness of, for example, about 200 to 1000 angstroms. For example, the SiGe layer 202 and the strained silicon layer 204 are formed by ultra high vacuum chemical vapor deposition (UHV-CVD) on the substrate. The SiGe layer 202 is epitaxially grown on the substrate

and the strained silicon layer 204 is epitaxially grown on the SiGe layer 202. For the UHV-CVD process, a reaction gas including Si_2H_6/GeH_4 is used, with a base pressure of about 2×10^{-10} Torr, a deposition pressure of less than 1 mTorr and a rising temperature gradient of 150°c/minute, for example.

[0026] Due to the lattice mismatch with the SiGe layer 202, the strained silicon layer 204 is under biaxial tensile strain. The band structure is modified and the carrier transport is enhanced. The tensile strain of the strained silicon layer 204 on the relaxed SiGe layer 202 also improves the so-called surface roughness scattering.

[0027] Referring to Fig. 2B, a gate oxide layer 206 is formed on the strained silicon layer 204. The gate oxide layer 206 is formed by thermal oxidation with a thickness of about 20-35 angstroms, for example.

[0028] It is noted that the underlying strained silicon layer 204 will be consumed and become thinner during the formation process of the gate oxide layer 206. The gate oxide layer 206 formed on the strained silicon layer 204 has a uniform thickness and the interface between the gate oxide layer 206 and the strained silicon layer 204 is smooth.

[0029] Afterwards, a gate structure 208 is formed on the gate

oxide layer 206. The material for forming the gate structure 208 is, for example, polysilicon or silicon germanium The process to form the gate structure 208 includes depositing a polysilicon layer or a silicon germanium layer (not shown) on the gate oxide layer 206 and patterning the polysilicon layer or the silicon germanium layer to form the gate structure 208, for example.

[0030] Referring to Fig. 2C, after forming the gate structure 208, the gate oxide layer 206 that is not covered by the gate structure 208 is removed, thus forming a gate oxide layer 206a remained underneath the gate structure 208.

[0031] Referring to Fig. 2D, a lightly doped drain (LDD) region 210 is formed in the strained silicon layer 204 at both sides of the gate structure 208. For example, once the gate structure 208 is established, the gate structure 208 serves to mask an lightly doped drain (LDD) implant into a certain thickness of the strained silicon layer 204 at both sides of the gate structure 208 to form the LDD region 210. Alternatively, the lightly doped region 210 is formed occupying the entire thickness of the strained silicon layer 204 at both sides of the gate structure 208 as shown in Figure 3A. The LLD implant comprises, for example, n-type species, such as, arsenic ions, for forming an NMOS

transistor. Further, the dosage and the implant energy are chosen such that the LDD region 210 is shallower than and of lighter concentration than the subsequent source/drain region.

[0032] Since the LDD region 210 is configured in the strained silicon layer 204 at both sides of the gate structure 208, a channel region 212 is defined in the strained silicon layer 204 between the LDD region 210.

[0033] Continuing to Figure 2E, a conformal spacer material (not shown) is deposited entirely over the semiconductor topography. The spacer material is, for example, a silicon nitride material or a silicon oxide material, and is preferably chemical vapor deposited (CVD) using the well known techniques. After the spacer material is deposited, a portion of the spacer material is removed to leave only the spacer material on the sidwall of the gate structure 208 as the spacer 214. The method to remove the portion of the spacer material includes anisotropic etching, such as, reactive ion etching.

[0034] Still referring to Figure 2E, an ion implantation process 216 is performed to implant dopants into the gate structure 208 and the strained silicon layer 204. The spacer 214 extends from the gate structure 208 a lateral distance

sufficient to mask the implants to form the heavily doped source/drain regions 218 in the strained silicon layer 204 and the silicon germanium layer 202 beside both sides of the gate structure 208. Alternatively, the heavily doped source/drain regions 218 are formed only in strained silicon layer 204 as shown in Figure 3B or only in the SiGe layer 202 as shown in Figure 3C. The ion implantation process 216 is, for example, a high concentration n-type implants, such as, an implant of arsenic ions with a dosage of about 3E15/cm^2 to 5E15/cm^2 and an implant energy of about 60KeV. Since n-type dopants, such as, arsenic ions, are implanted, after the ion implantation process 216, the gate structure 208 is transformed into an n-type gate structure 208a.

[0035] Continuing to Figure 2F, a cap layer 220 is deposited entirely over the semiconductor topography, covering the gate structure 208a, the spacer 214 and the surface of the source/drain regions 218. The cap layer 220 is, for example, a silicon oxide layer formed to a thickness of about 300 angstroms to about 700 angstroms. The cap layer 220 is formed by, for example, a chemical vapor deposition technique at a temperature of about 400 degrees Celsius to about 500 degrees Celsius. After the cap layer

230 is deposited, a rapid thermal annealing (RTA) process is then conducted. The RTA process is conducted, for example, at a temperature about 1000 degrees Celsius for a duration of about 10 to 20 seconds. Due to the cap-annealing process, a compressive stress is generated on the n-type gate structure 208a. Consequently, the biaxial tensile strain in the strained silicon layer 204 underneath the gate structure 208a is enhanced.

[0036] Referring to Figure 2G, the cap layer 220 is subsequently removed. Thereafter, a metal layer (not shown) is formed over the semiconductor topography, covering the gate structure 208a, the spacer 214 and the surface of the heavily doped source/drain regions 218. The metal layer, which includes W, Ti, Ni, Mn, Co, Pt or Pd is formed by, for example, sputtering. After executing an annealing process, for example, at a temperature of about 400 to 800 degrees Celsius for about 20 to 60 seconds, the excess unreacted metal layer is selectively removed, for example, by wet etching. A silicide film 234 is formed atop the gate structure 208a and the heavily doped source/drain regions 218 to lower the sheet resistance at the source/drain area and the gate electrode. In the case a nickel silicide is formed, the consumption of silicon in the gate

structure 208a and the strained silicon layer 204 can also be reduced. In the case the gate structure 208 is formed with a silicon germanium layer, the gate resistance can be further reduced. Similarly, when the heavily doped source/drain region is configured in the strained silicon layer 204 and the silicon germanium layer 202, a very low sheet resistance silicide, constructed with silicon germanium, can also be formed in the source/drain area.

[0037] Due to lattice mismatch with the SiGe layer 202, the strained silicon layer 204 is under biaxial tensile strain when it is being epitaxially grown on a SiGe layer. During the cap annealing process, a compressive stress is generated on the n-type gate structure 208bs due to the expansion of the arsenic ions implanted in the gate structure 208. A tensile strain is thereby further introduced to the strained-channel 212. The induced tensile strain modifies the band structure and enhances the carrier transport. Thus, if a field effect transistor is formed by using the strained silicon layer 204 as the channel, the horizontal mobility of electrons increases in the channel, and the operating speed of the transistor also increases. A higher amount of strain can be induced from generating tensile stress in the n-type gate.

[0038] The tensile strain of the strained silicon layer 204 on the relaxed SiGe layer 202 improves the so-called surface roughness scattering and the phonon scattering. Due to the improved surface roughness scattering-limited mobility and the phonon scattering-limited mobility of strained silicon, the electron mobility increases. Further, the current drive characteristics of a strained silicon device also exhibit improvement.

[0039] In accordance to the present invention, the source/drain region is formed in the strained silicon layer, the channel region is thus also located in the strained silicon layer. Because the conductivity of the SiGe layer is higher and the electron mobility in the strained silicon layer is faster compared with the conventional silicon substrate, resistance of the source/drain region can be efficiently reduced. Therefore, the source/drain region of the present invention can adopt shallow junctions, thus avoiding short channel effect and junction leakage. As a result, not only the device performance but also reliability of the device can be improved. Accordingly, the strained Si MOSFET of the present invention is promising in high performance deep sub-100nm (sub-0.1micron) CMOS technology.

[0040] In conclusion, in accordance to the manufacturing method

of a semiconductor device of the present invention, a compressive stress is generated on the N-type gate. A tensile strain is also further being introduced to the channel disposed in the strained silicon layer underneath the N-type gate. The horizontal mobility of electrons thus increases in the channel.

[0041] Further, the source/drain region is located either in the strained silicon layer or in both the strained silicon layer and the SiGe layer with higher conductivity, the resistance of the source/drain region is effectively reduced.

[0042] Since the resistance of the source/drain region is effectively reduced, the source/drain region can be formed as shallow junctions, thus avoiding short channel effect and junction leakage.

[0043] For the manufacture method of a semiconductor device of the present invention, both the performance and the reliability of the device are improved.

[0044] Although in the above, the present invention has been described with respect to a NMOS device, the present invention is also applicable to devices, such as, high/low voltage transistors in the peripheral device region or other types of transistors.

[0045] It will be apparent to those skilled in the art that various

modifications and variations can be made to the structure of the present invention without departing from the scope or spirit of the invention. In view of the foregoing, it is intended that the present invention cover modifications and variations of this invention provided they fall within the scope of the following claims and their equivalents.